OPEN ACCESS

University of BRISTOL

Basci, F., & Kocak, T. (2004). A low-power network search engine based on statistical partitioning. In *IEEE Workshop on High Performance Switching and Routing, Phoenix AZ* (pp. 264 - 268). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/HPSR.2004.1303484

Peer reviewed version

Link to published version (if available):
10.1109/HPSR.2004.1303484

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

A Low-Power Network Search Engine Based on Statistical Partitioning

Taskin Kocak and Faysal Basci
Department of Electrical and Computer Engineering
University of Central Florida
Orlando, Florida 32816-2450
E-mail : tkocak@cpe.ucf.edu; tel: 407-823-4758; fax: 407-823-5835

Abstract—**Network search engines based on Ternary CAMs are widely used in routers. However, due to parallel search nature of TCAMs power consumption becomes a critical issue. In this work we propose an architecture that partitions the lookup table into multiple TCAM chips based on individual TCAM cell status and achieves lower power figures.**

I. Introduction

Content addressable memory (CAM) provides access by data rather than by memory address. CAMs have higher advantage over other memory search algorithms, such as look-aside tag buffers, binary or tree based searches. However, this performance advantage comes with a price of higher silicon area, and higher power consumption. Today, commercial CAM chips or embedded CAMs are being utilized in lots of different applications including pattern recognition, neural networks, encryption, firewalls, switches and routers. However, in this paper, we are interested in the ones for networking applications. CAMs are generally used in packet forwarding lookup tables in the routers [1]. It is used to extract and process the address information from incoming packets: compare the destination address of the packet with the stored data and if a match occurs, associated routing information is given to the forwarding circuit. Despite their performance advantages, CAMs have serious power consumption problems. In [1], it is reported that a 500K-entry lookup table for an IPv6 network processor, formed by CAM chips will consume up to 133 W, which is around 3 W per chip. Moreover, in [7] it is reported that a 64K-word by 40-bit CAM chip consumes 5.2 W.

In this paper, we are going to address the power consumption issues in ternary CAMs (TCAMs) used in IP forwarding tables and propose an approach which reduces the expected power consumption. In section II, an introduction to fundamentals of CAMs is presented, section III discusses usage of TCAMs in IP forwarding, and section IV presents our statistical partitioning approach. In section V, power consumption formulation of statistically partitioned TCAM is presented, and section VI discusses the experimental results. Finally, in section VII, we present the conclusion.

II. CAM Basics

Fig. 1 shows a basic CAM architecture [3]. This is an n-bit k-word CAM. Data stored in CAM are searched by applying the reference word to bit lines, which run vertically, through bit line drivers. When a cell detects a mismatch, it will pull down the match line it is connected to. If a word is fully matched the match line will remain high. The encoder will select a single row in the case of multiple matches, and will assert a hit signal and the corresponding address of the selected row. As can be seen from the figure any search word presented is searched in parallel. Due to this, the search is very fast, however this also implies that for each search operation all of the cells are utilized, resulting in excessive power consumption.

There are two classes of CAMs, binary CAMs and ternary CAMs. A binary CAM cell can store either a ***0*** or a ***1***. TCAMs, on the other hand, has the capability to store a "don't care (x)". To store an x we need an extra bit. This can be achieved by simply combining two binary CAM cells [4] or a more elaborate CAM cell design is also possible as in [9]. In TCAM cells the stored data is encoded to represent ***0***, ***1*** and x. CAM architectures utilizes wire-anding; before a search operation is performed, the matchline is precharged and search lines are discharged. During a search operation, if a cell matches the stored data with the one on the search line it will do nothing, however when a mismatch is detected, the cell will pull down the match line to low. So, even if one bit mismatches, a mismatch will be issued. On the other hand, if a don't care is stored in a cell then the search data will be discarded by that cell. In a way it will behave like a matching cell.

CAMs (or TCAMs) consume power mainly in 3 parts: matchline (and searchline) precharging (pre-discharging), comparison, and clock and control signaling. Most of the power is consumed during pre-charging operation [6]. Therefore most of the CAM designs try to minimize pre-charging events as in [5], [8] and [10]. Some approaches try to reduce the voltage swing across search and matchlines as in [11] whereas some approaches use system level optimizations. For example in [12], a parameter characterizing the data is also stored in the CAM besides the original data. First, a search among the parameters is performed and a second search is done among the original data for which corresponding parameter search returned a hit. This way evaluation and precharging activities are avoided.

Other system level approaches utilizes the application specific properties, as in the case of IP forwarding engines. Next section discusses the usage of TCAMs in IP forwarding.

0-7803-8375-3/04/$20.00 © 2004 IEEE

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on July 9, 2009 at 06:10 from IEEE Xplore. Restrictions apply.

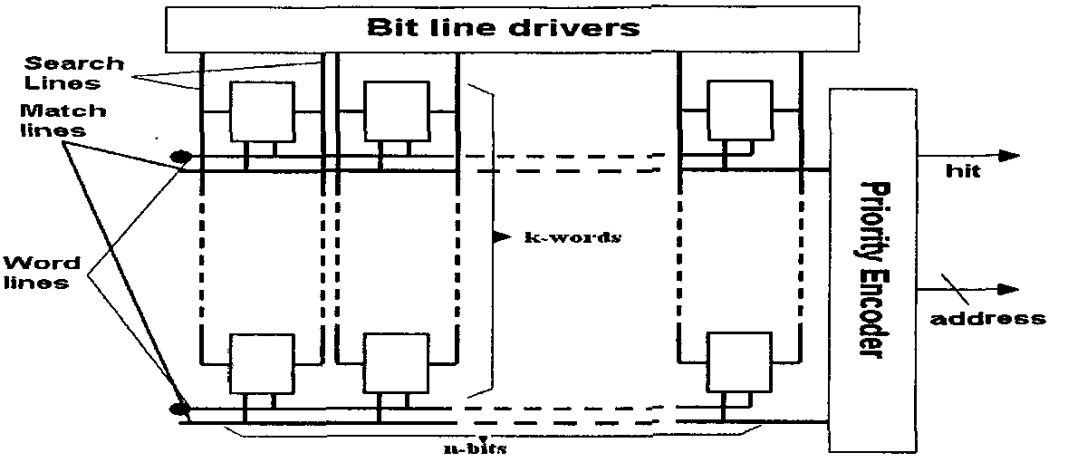

Fig. 1. CAM architecture. Data to be searched (or stored) is fed through bit line drivers, address is the encoded address of the matched data. Hit is a control signal indicating if a match found. Word lines are used when writing data to the cells.

III. TCAMs in IP forwarding

Nowadays, IP lookups are based on classless interdomain routing (CIDR) scheme. After the adoption of CIDR in 1993, IP routes have been characterized by a routing prefix and the prefix length. In CIDR scheme, route lookups aim at the longest prefix match (LPM).

As a TCAM providea a don't care storage capability, it is a favorable lookup hardware; in that, in IP lookup engines, when an entry is stored in a TCAM, depending on its prefix, some of its rightmost bits will be stored as x. For example in IPv4, for a 24 bit prefix, last 8 bits will be x. Moreover, because of x storage capability, routing entries belong to different prefix sets can be placed on the same chip. When multiple matches occur, a priority encoder chooses the longest matching prefix. In the case of binary CAMs, one needs to use a separate CAM chip for each prefix set.

Routing table entries stored in TCAMs are ordered according to their prefixes. For example in Fig. 2, the highest prefix set lies at the top of the entries (or lowest addresses), and the lowest prefix set lies at the bottom [14]. So in the case of multiple matches priority encoder will choose the one with lowest address, which is indeed the longest matching prefix.

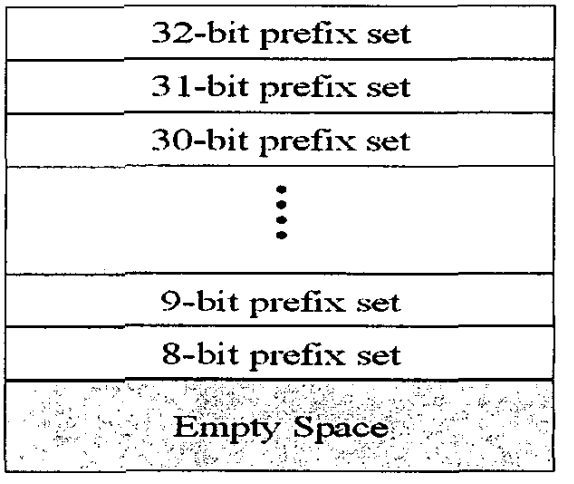

Fig. 2. TCAM storage scheme. Higher prefix sets are at top, lower prefix sets are at the bottom.

IV. Statistical Partitioning

When we look at the prefix distribution in the core routers we see that they all have a similar characteristic. Fig. 3 shows the distribution of prefixes in the routing tables that belong to routers of various networks [13]. It can be seen that a great portion of the entries are accumulated at prefix set 24. The question is can we make use of this? At first, If we just check this prefix set we can achieve around 50% hit ratio provided that we have a random traffic pattern. However, as our aim is to find the longest matching prefix, we should also look at the higher prefixes. Fig. 3 suggests that prefixes higher then 24 are considerably rare. So, we exploit this fact and propose to partition the routing lookup table as shown in Fig.4. In this architecture, the routing lookup table is divided into two parts, TCAM1 and TCAM2, respectively. In a lookup operation, first TCAM1 will be searched and if a mismatch occurs, then TCAM2 will be searched. The search procedure can be summarized as follows:

Perform a search in TCAM1
  If there is a match
    DONE
  Otherwise
    search TCAM2

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on July 9, 2009 at 06:10 from IEEE Xplore. Restrictions apply.

If there is a match
DONE
Otherwise
issue mismatch
end

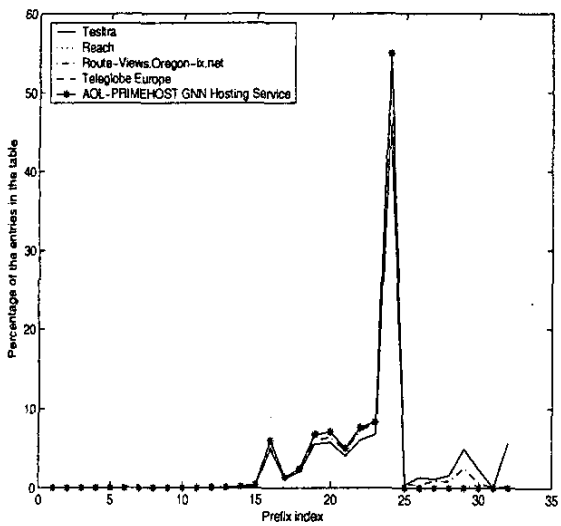

Fig. 3. Prefix distribution in the routing tables of various networks

There is a buffer between TCAM1 and TCAM2 which is used to store the current search word. This enables pipelining the search operation. In the case of a mismatch in TCAM1, TCAM2 uses the value stored in the buffer to do a search. This way, TCAM1 can accept a new search word each time. Although pipelining ensure a sustained average throughput of one result per clock cycle, the average latency of the search operation will be higher. However, considering the overall latency of packet processing in network nodes, we believe that this latency will not be significant as long as the throughput rate is sustained.

Beside the distribution of prefixes we should also take power consumption and average latency into account when partioning the table. When we have all of these figures we can deduce an optimal partitioning.

V. Power Consumption Formulation of CAM Circuits

Power consumption in a TCAM cell can be written as follows:

$$P_{TCAM} = P_{STC} + P_{CLK} + P_{MISS} + P_{MATCH} + P_X \quad (1)$$

where P_{STC}, P_{CLK} represent the static power consumption and power dissipation due the clock circuitry, and P_{MISS}, P_{MATCH}, P_X represent the average power consumption in the case of a mismatch, match and don't care, respectively. Please note that latter three also include the power consumption due to matchline and searchline switchings.

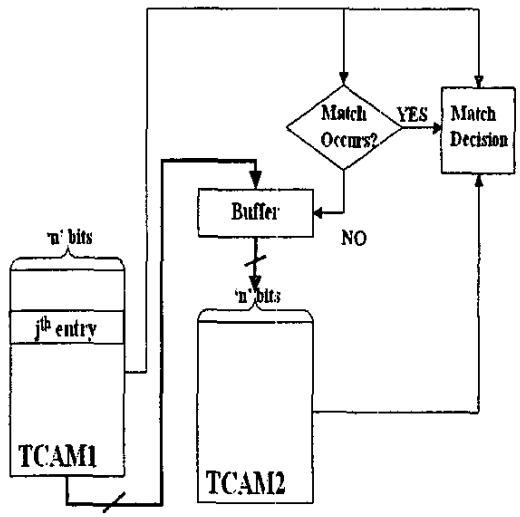

Fig. 4. Partitioned TCAM

Let r_i represent the set of entries associated with a prefix i, and prefix sets stored in TCAM1 be $\{r_m, \ldots r_n\}$ where m < n, and prefix sets stored in TCAM2 be $\{r_1, r_2, \ldots r_{m-1}\}$. The number of entries in set r_i can be represented by N_i, where 1< i <n . The number of don't care cells in a prefix i is n-i. Then power consumption for a comparison operation of a word that belongs to the set r_i is:

$$P^i_{COMP_j} = P_{MATCH} * Mt^i_j + P_{MISS} * Ms^i_j + P_X * (n-i) \quad (2)$$

where Mt^i_j represents the number of matching cells and Ms^i_j represents the number of mismatching cells in stored word j. Let ρ represent the probability that a match will occur in TCAM1, then the total power consumption for a search operation can be written as:

$$P_{TOTAL} = P_{TCAM1} + (1 - \rho_o * \rho) * P_{TCAM2} \quad (3)$$

where ρ_o represents the probability that the search word has a match in the whole table. Then we can write the power consumptions for each TCAM portion as:

$$P_{TCAM1} = \sum_{i=m}^{n} \sum_{j=1}^{N_i} P^i_{COMP_j} \quad (4)$$

$$P_{TCAM2} = \sum_{i=1}^{m-1} \sum_{j=1}^{N_i} P^i_{COMP_j} \quad (5)$$

To proceed further, note that

$$0 \leq Mt^i_j \leq i \quad (6)$$

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on July 9, 2009 at 06:10 from IEEE Xplore. Restrictions apply.

$$0 \leq Ms_j^i \leq i \tag{7}$$

Assuming that the number of mismatching and matching cells are equal (contradicting case will be discussed later)

$$E\left\{Mt_j^i\right\} = i/2; \tag{8}$$

$$E\left\{Ms_j^i\right\} = i/2; \tag{9}$$

where $E\{...\}$ represents the expected value operation. Then the expected value of power consumption for each TCAM reduces to the following:

$$E\{P_{TCAM1}\} = \tag{10}$$
$$\sum_{i=m}^{n} N_i * [P_{MATCH} \cdot i/2 + P_{MISS} * i/2 + P_X * (n-i)]$$

$$E\{P_{TCAM2}\} = \tag{11}$$
$$\sum_{i=1}^{m-1} N_i [P_{MATCH} * i/2 + P_{MISS} * i/2 + P_X * (n-i)]$$

Hence we can write the expected total power expression as follows:

$$E\{P_{TOTAL}\} = E\{P_{TCAM1}\} + (1 - \rho_o * \rho_1) * E\{P_{TCAM2}\} \tag{12}$$

VI. Experimental results

After formulating power, the problem is to find the value of m that will minimize the expected total power consumption, while maintaining an acceptable average search latency. As the value of m depends on the routing table entries and traffic pattern one would expect that it will be different for each lookup table. To calculate the expected power consumption for different values of m, or a typical lookup table, we have done calculations for a few networks including Telstra and Reach [13]. We have implemented an example TCAM circuit in TSMC $0.18\mu m$ CMOS technology and we run it at 100 MHz. P_{MT}, P_{MS} and P_X are obtained from Cadence Spectre simulations. In these simulations, number of TCAM cells varied and in each case TCAM block is tested with different matching and mismatching bit patterns. Finally, all of the obtained results are averaged for matching, miss-matching and don't care cells. The results are as follows:

$$P_{MT} = 397\ nW, P_{MS} = 515\ nW, P_X = 336\ nW \tag{13}$$

Fig. 5, shows the power consumption in search operations for 5 different networks, based on the above values. It can be seen that for all of the networks the minimum power consumption is achieved for the value of m=24. Power savings gained by doing that are between 20 to 24%. We can also use expected power - latency product (PLP) as another metric. If we represent search latency in terms of clock cycles with L,

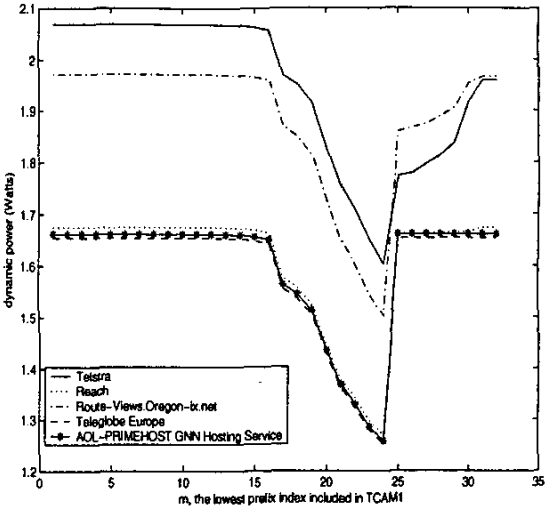

Fig. 5. Power consumption in search operation for partitioned CAM with varying m

Network	m1, m2	Power saving (%)	Average latency (cc)
Telstra	25, 24	30.23	1.6754
Reach	24, 21	28.65	1.6871
Route-Views.Oregon	24, 21	27.62	1.6321
Teleglobe Europe	24, 21	28.89	1.6932
AOL-PR. GNN	24, 21	28.84	1.6914

TABLE I
EXPECTED POWER SAVINGS AND AVERAGE LATENCY WHEN PARTITIONING TCAM INTO THREE

expected value of the latency can be shown to be equal to the following :

$$E\{L\} = 2 - \rho_1 \tag{14}$$

where ρ_1 is the probability that a match will occur in TCAM1. Fig. 6 shows the plot of PLP for the same network set. It can be seen that the until m=25 the PLP increases very slowly, and then it shows a sharp increase. When we calculate average latency for the case where m=24, we see that it varies from 1.36 to 1.45 clock cycles. The same experiment is run for the case of 3 partitions. This time, we have 3 TCAM parts. The prefix set stored in TCAM1 in this case is $\{r_{m_1}, \ldots r_n\}$ m_1 < n, prefix set stored in TCAM2 is $\{r_{m_2}, \ldots r_{m_1-1}\}$ $m_2 < m_1$ and prefix set stored in TCAM3 is $\{r_1, \ldots r_{m_2-1}\}$. After running the simulations based on this scheme, the obtained minimum power values are tabulated in Table I. From the table it can be seen that with some added latency, power consumption can be reduced by up to 30%. Experiments we conducted show that further partitioning does not increase power savings considerably.

In the experiments, it is assumed that the number of cells that return a match is equal to the number of cells returning a mismatch. However, to exemplify the power consumption figure for different miss rates a simulation is done for Telstra

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on July 9, 2009 at 06:10 from IEEE Xplore. Restrictions apply.

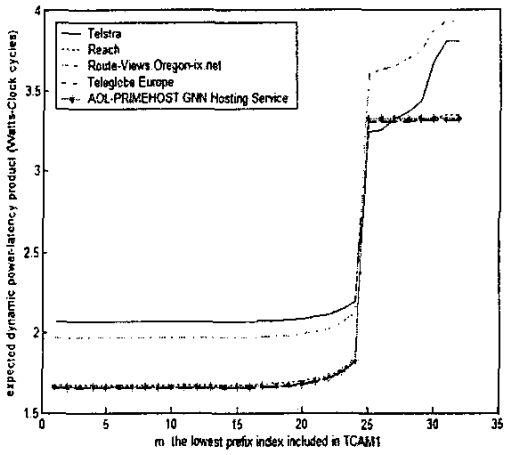

Fig. 6. Expected power-latency product (PLP) with varying m

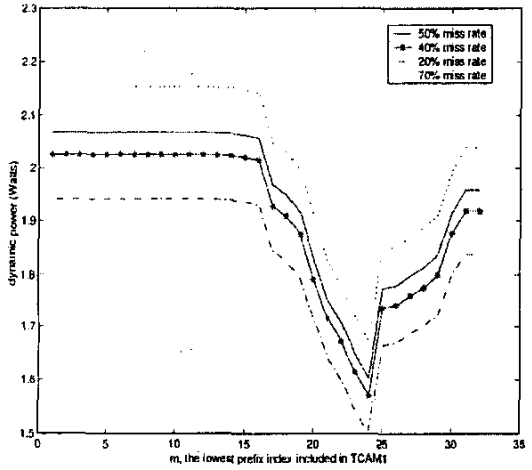

Fig. 7. Power consumption in search operation for partitioned CAM with varying m for Telstra Network with different miss rates for individual cells

network. The result is shown in Figure 7. As can be seen the optimal value of m is not affected, however the power consumption figures are scaled up or down with miss rate.

VII. Conclusion

We have presented a partitioning scheme, which utilizes statistical distribution of prefixes and individual power consumption of cells in the cases of match, mismatch and don't cares. We showed that indeed the partitioning helps reducing the power consumption in IP lookup applications. Partitioning into two and three, reduce power consumption considerably, whereas further partitioning shows little improvement.

References

[1] White Paper, EZCHIP Technologies, "Ipv6 to Ipv4 is Not Merely 50 More." (web: http://www.ezchip.com/html/tech IPv6.html)

[2] "Content-Addressable memory (CAM) and its network applications." (web: www.eetasia.com/ARTICLES/2000MAY/2000MAY03_MEM_NTEK_TAC.PDF)

[3] K. J. Schultz, "Content-addressable memory core cells A survey." *Integration, the VLSI journal 23*, Page(s): 171-188, 1997

[4] R. Sergio, R. Chavez, "Encoding don't cares in static and dynamic Content-Addressable Memories." *IEEE Transactions on Circuits and Systems-II : Analog and Digital Signal Processing* Vol. 39, No.8, August 1992

[5] G. Thirugnanam, N. Vijaykrishnan, M.J. Irwin, "A novel low power CAM design." *14th Annual IEEE International ASIC/SOC Conference Proceedings*, Page(s): 198 -202, 12-15 Sept. 2001

[6] H.Y. Liang Hsiao; D.H. Wang; C.W. Jen, "Power modeling and low-power design of content addressable memories " *ISCAS, The 2001 IEEE International Symposium on Circuits and Systems*, Volume: 4 , Page(s): 926 -929 vol. 4, 6-9 May 2001

[7] F. Hafai, K.J. Schultz, G.F.R. Gibson, A.G. Bluschke, D.E. Somppi, "Fully parallel 30-MHz, 2.5-Mb CAM." *IEEE Journal of Solid-State Circuits*, Volume: 33 Issue: 11, Page(s): 1690 -1696, Nov. 1998

[8] C.A. Zukowski, S.Y. Wang, "Use of selective precharge for low-power on the match lines of content addressable memories." *International Workshop on Memory Technology, Proceedings of Design and Testing*, Page(s): 64 -68, 11-12 Aug. 1997

[9] I. Arsovski, T. Chandler, A. Sheikholeslami, "A ternary content-addressable memory (TCAM) based on 4T static storage and including a current-race sensing scheme" *IEEE Journal of Solid-State Circuits*, Volume: 38 Issue: 1, Page(s): 155 -158, Jan. 2003

[10] T. Chadwick, T. Gordon, R. Nadkarni, J. Rowland, "An ASIC-embedded content addressable memory with power-saving and design for test features." *IEEE Conference on Custom Integrated Circuits*, Page(s): 183 -186, 6-9 May 2001

[11] H. Miyatake, M. Tanaka, Y. Mori, "A design for high-speed low-power CMOS fully parallel content addressable memory macros." *IEEE Journal of Solid- State Circuits*, Volume: 36 Issue: 6, Page(s): 956 -968, June 2001

[12] C.S. Lin, J.C. Chang, B.D. Liu,"A low-power precomputation-based fully parallel content-addressable memory." *IEEE Journal of Solid-State Circuits*, Volume: 38 Issue: 4, Page(s): 654 -662, April 2003

[13] BGP table statistics, (web: http://bgp.potaroo.net/), September 22, 2003

[14] D.Shah, P.Gupta, "Fast updating algorithms for TCAM" *IEEE Micro*, Volume: 21 Issue: 1, Page(s): 36 -47, Jan.-Feb. 2001

Authorized licensed use limited to: UNIVERSITY OF BRISTOL. Downloaded on July 9, 2009 at 06:10 from IEEE Xplore. Restrictions apply.